Yeonji Jung
Madison Volpe

Scraping Boston Crime Data

A) Create a tibble called crime.data which contains *all* crimes in *all* neighborhoods (i.e., each row represents a crime), and has three columns: -crime (the name of the crime, from the 'Type' field on each page) -hour (the hour as an integer from 0 to 23, from the 'Date' field) -neighborhood (the neighborhood name as a string)

```
# A tibble: 6 x 3
   neighborhood crime              hour
   <chr>        <chr>             <int>
1  allston      rape                  3
2  allston      shooting             22
3  allston      bank robbery          5
4  allston      stabbing              1
5  allston      armed robbery        22
6  allston      bank robbery         11

Classes 'tbl_df', 'tbl' and 'data.frame':      1128 obs. of  3 variables:
 $ neighborhood: chr  "allston" "allston" "allston" "allston" ...
 $ crime       : chr  "rape" "shooting" "bank robbery" "stabbing" ...
 $ hour        : int  3 22 5 1 22 11 2 23 2 0 ...
```

B) What are the five most common crime types (aggregated across neighborhoods and hours), and how many of each such crime occurred? Be alert for misspellings!

The five most common crime types (aggregated across neighborhoods and hours) are:

- o **Shooting – 227 occurrences**
- o **Gunfire – 143 occurrences**
- o **Murder – 141 occurrences**
- o **Assault – 132 occurrences**
- o **Stabbing – 119 occurrences**

Yeonji Jung
Madison Volpe

C) Make a plot of the total number of crimes (aggregated across neighborhoods and crime types) by hour. Write a few sentences about the pattern you observe.

Total Number of Crimes across Boston Neighborhoods by Hour
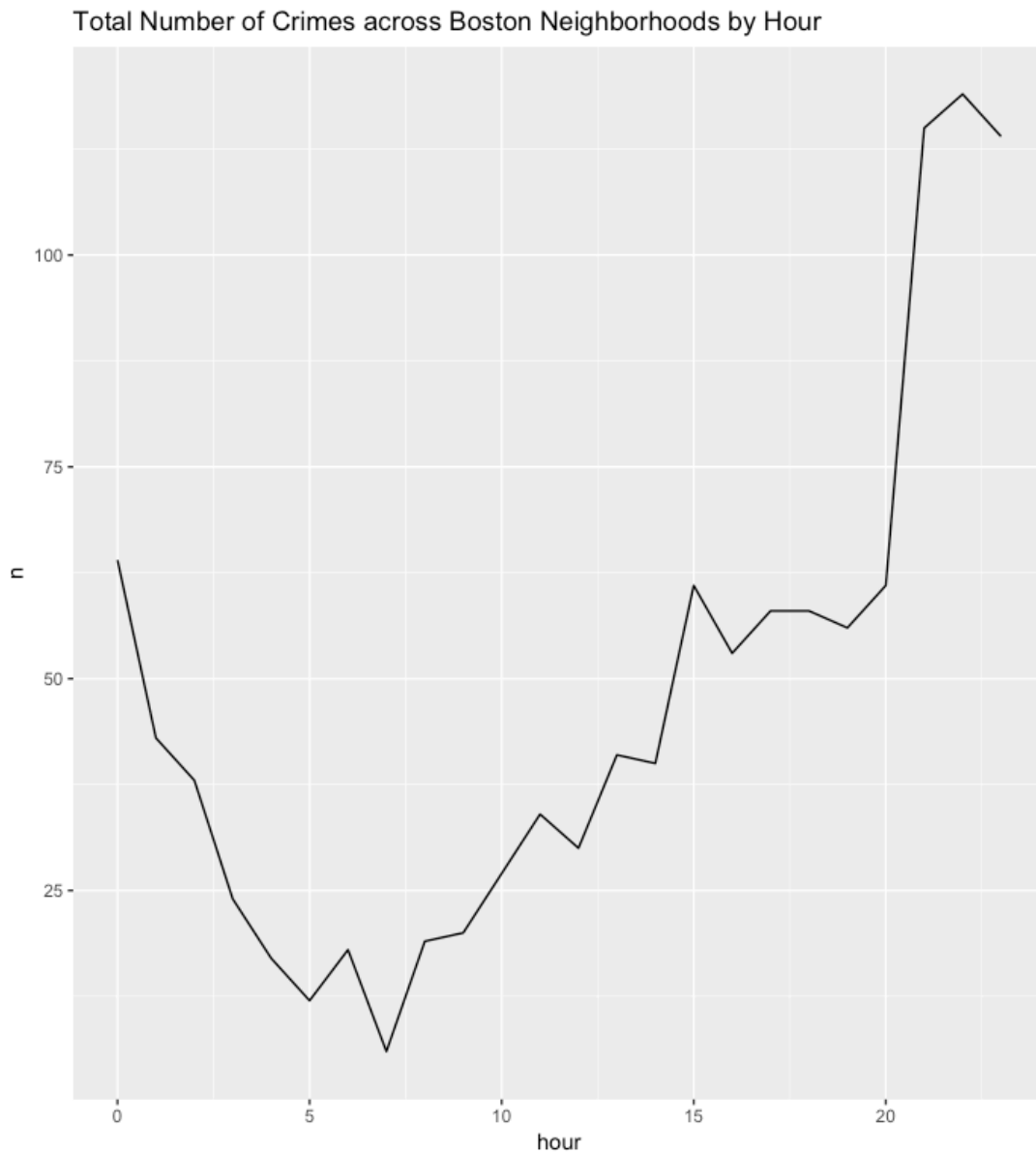
This line plot allows us to observe a variety of patterns. The most obvious is that crimes tend to be committed during later hours, such 9 pm (21), 10 pm (22), 11 pm (23), and 12 am (0). Crimes tend to be less frequently committed during typical work/daytime hours, such as 5 am (5), 6 am (6), 7 am (7) and so on.

D) Restrict to the five most common crime types, and plot the total number of crimes (aggregated across neighborhoods) for each crime type by hour (i.e., your plot should have five lines). Write a few sentences about the pattern you observe.
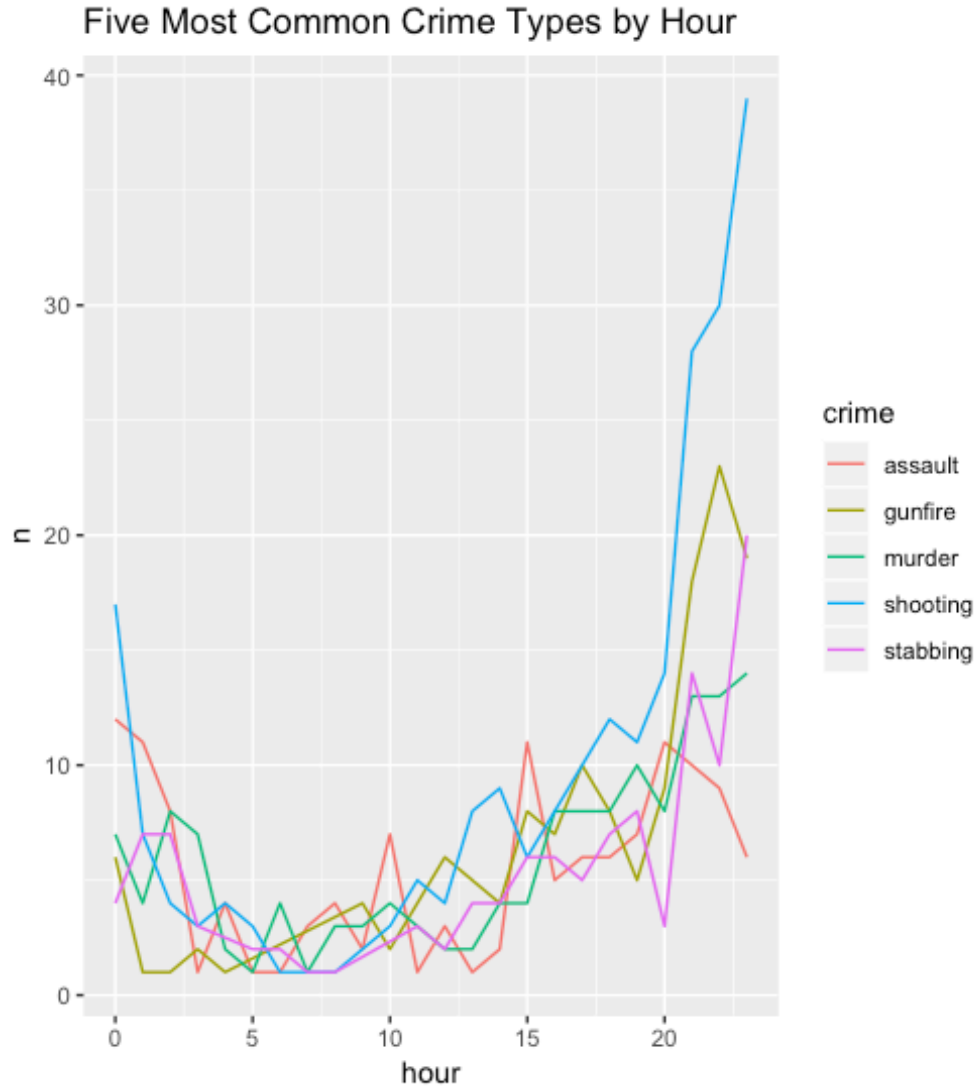
From the line graph, we see that shootings are by far the most frequent crime type from 8pm (20) until 11pm (23). At 12 am (0) shootings remain the most frequent crime type, however they decrease in frequency in typical working and commuting hours. We can also see that gunfire is correlated with shootings, as the lines that represent these crimes tend to increase and decrease at similar points. Assaults appear to be the most frequent crime during typical working and commuting hours.

Yeonji Jung
Madison Volpe

E) Restrict to just the neighborhoods of Dorchester and Downtown, and plot the total number of crimes (aggregated across crime types (include all crime types, not just the top five)) for each of the two neighborhoods by hour (i.e., your plot should have two lines). Write a few sentences about the pattern you observe.

Dorchester and Downtown Total Number of Crimes by Hour
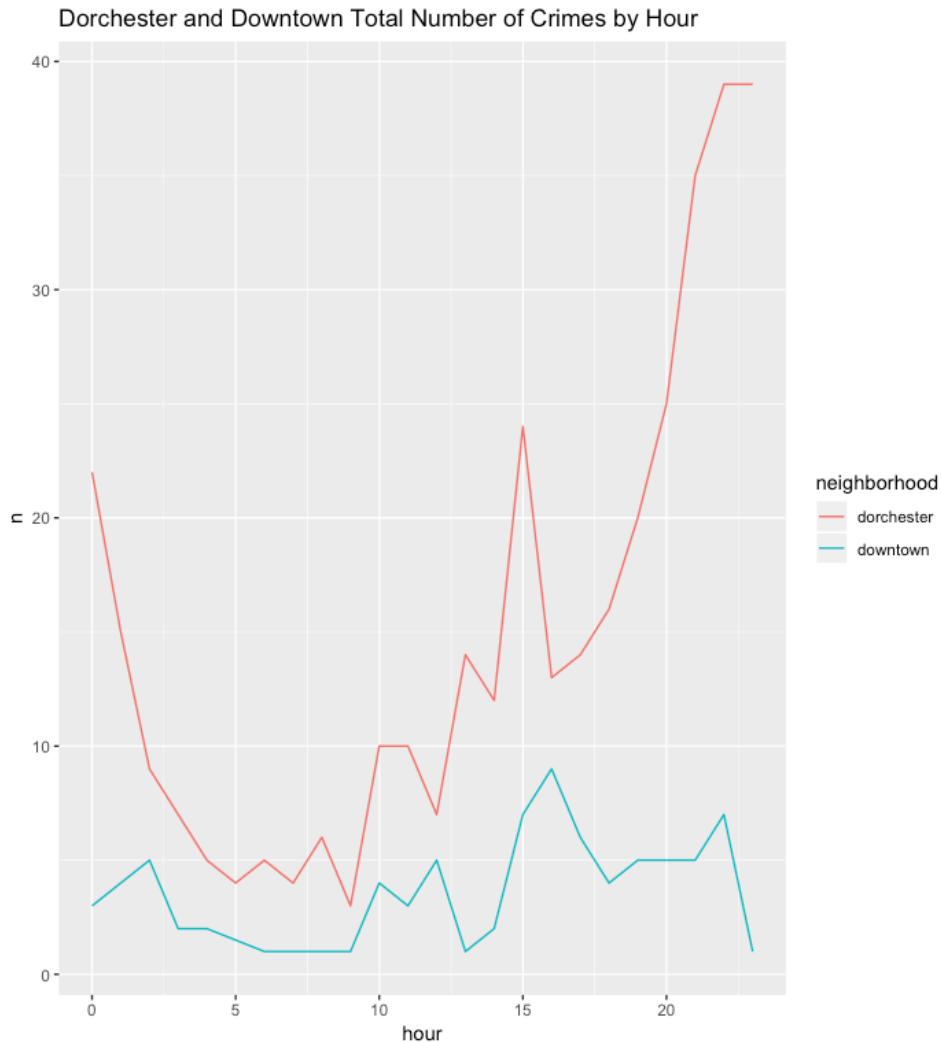
From the above plot, we can see that no matter the hour, Dorchester tends to have the higher total number of crimes compared to Downtown. For both Dorchester and Downtown, there appears to be a spike in crime at around 3 pm (15), this then sharply declines for Dorchester and slightly declines for Downtown. The total number of crimes then steadily increases for Dorchester from 4 pm (16) to 11 pm (23). Following a slight spike in crime at around 9 or 10 pm (21/22), crime steadily declines Downtown at 11 pm (23).